

CC -1- FUNCTION: NSB IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M67463; AAA5534.1; -
 CC PIR: A36814; GNMVCH
 CC PDB: 1HEI; 25-NOV-98.
 CC PDB: 1AIV; 16-FEB-99.
 CC
 CC MEROPS: S29.001; -
 CC
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS4c.
 CC InterPro: IPR002166; HCV_NS4d.
 CC InterPro: IPR002522; HCV_NS4e.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_core.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; HCV_NS5a; 1.
 CC Pfam: PF00271; Helicase_C; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00492; Helicase3; 1.
 CC
 CC Polyprotein: Glycoprotein: Transferase: RNA-directed RNA polymerase;
 CC Core protein: Coat protein: Envelope protein: Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein: Hydrolase; Serine protease;
 CC 3D-structure.
 CC
 CC INIT_MET 1 1
 CC
 CC CHAIN 1 191
 CC CHAIN 192 383
 CC CHAIN 384 746
 CC CHAIN 747 809
 CC CHAIN 810 1026
 CC CHAIN 1027 1657
 CC CHAIN 1658 1711
 CC CHAIN 1712 1972
 CC CHAIN 1973 2420
 CC CHAIN 2421 3011
 CC TRANSMEM 347 369
 CC ACT_SITE 1083 1083
 CC ACT_SITE 1107 1107
 CC ACT_SITE 1165 1165
 CC NP_BIND 1230 1237

REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.

CAPSID PROTEIN C.

ENVELOPE GLYCOPROTEIN E1.

ENVELOPE GLYCOPROTEIN E2.

PROTEIN P7.

NONSTRUCTURAL PROTEIN NS2.

PROTEASE/HELICASE NS3.

NONSTRUCTURAL PROTEIN NS4a.

NONSTRUCTURAL PROTEIN NS4b.

NONSTRUCTURAL PROTEIN NS5a.

NONSTRUCTURAL PROTEIN NS5b.

POTENTIAL.

CHARGE RELAY SYSTEM (BY SIMILARITY).

CHARGE RELAY SYSTEM (BY SIMILARITY).

CHARGE RELAY SYSTEM (BY SIMILARITY).

ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CC094753 CRC64;

Query Match 95.2%; Score 2395; DB 1; Length 3011;
 Best Local Similarity 97.0%; Pred. No. 3,9e-169;
 Matches 452; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 2 VDFIPVENLETTMRSPVFTDSSPPVYPOSFQVAAHAPRGSKSTKVPVAAAGYKVL 61
 DB 1193 VDFIPVENLETTMRSPVFTDSSPPVYPOSFQVAAHAPRGSKSTKVPVAAAGYKVL 1252
 QY 62 VLNPSVAATLGFAGVMSKAGIDPNRTGVTYTTGSPITYSTGKFLADGGCGGAYVDI 121
 DB 1253 VLNPSVAATLGFAGVMSKAGIDPNRTGVTYTTGSPITYSTGKFLADGGCGGAYVDI 1312
 QY 122 IICDECHSTDATSIIGIGTVLDOAETAGARLVLAATPPGSAVTPPHNIEVALSTTGE 181
 DB 1313 IICDECHSTDATSIIGIGTVLDOAETAGARLVLAATPPGSAVTPPHNIEVALSTTGE 1372
 QY 182 IPEYGAIPLEVIKGGHLLFCHSKKKCDELAALVALGINAAVAYVRLDVSVPISGDV 241
 DB 1373 IPEYGAIPLEVIKGGHLLFCHSKKKCDELAALVALGINAAVAYVRLDVSVPISGDV 1432
 QY 242 VVSTDAIMGTFDPSVVCNTCWTVDSSLDPTTSTTLPDAAVSTQRRGRGTG 301
 DB 1433 VVSTDAIMGTFDPSVVCNTCWTVDSSLDPTTSTTLPDAAVSTQRRGRGTG 1492
 QY 302 RGKGIYRFVAPGERPSGMPDSSVLCEDYDAGCAVETLPAETTVRLRAYNTPGLPVCO 361
 DB 1493 RGKGIYRFVAPGERPSGMPDSSVLCEDYDAGCAVETLPAETTVRLRAYNTPGLPVCO 1552
 QY 362 DHLEFMEGVETGLHIDAHFISQTRQSGENLPYLAAYQATVCAARAQAPPSDDOMKCL 421
 DB 1553 DHLEFMEGVETGLHIDAHFISQTRQSGENLPYLAAYQATVCAARAQAPPSDDOMKCL 1612
 QY 422 RLKPTLHGPTPLLYRLGAVONEITLTHPTKTIIMTCSADLEVTYS 467
 DB 1613 RLKPTLHGPTPLLYRLGAVONEITLTHPTKTIIMTCSADLEVTYS 1658

RESULT 2
 POLG_HCVH
 ID POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.